Convex Optimization M2

Lecture 3

Duality

DMs

DM par email: dm.daspremont@gmail.com

Outline

- Lagrange dual problem
- weak and strong duality
- geometric interpretation
- optimality conditions
- perturbation and sensitivity analysis
- examples
- theorems of alternatives
- generalized inequalities

Lagrangian

standard form problem (not necessarily convex)

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

variable $x \in \mathbb{R}^n$, domain $\mathcal{D}$, optimal value $p^\star$

Lagrangian: $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$, with $\mathbf{dom}\, L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$,

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $f_i(x) \leq 0$
- ν_i is Lagrange multiplier associated with $h_i(x) = 0$

Lagrange dual function

Lagrange dual function: $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$,

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \\ &= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right) \end{aligned}$$

g is concave, can be $-\infty$ for some λ, ν

lower bound property: if $\lambda \succeq 0$, then $g(\lambda, \nu) \leq p^\star$

proof: if $\tilde{x}$ is feasible and $\lambda \succeq 0$, then

$$f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu)$$

minimizing over all feasible $\tilde{x}$ gives $p^\star \geq g(\lambda, \nu)$

Least-norm solution of linear equations

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b \end{array}$$

dual function

- Lagrangian is $L(x,\nu) = x^T x + \nu^T(Ax - b)$
- to minimize L over x, set gradient equal to zero:

$$\nabla_x L(x,\nu) = 2x + A^T\nu = 0 \quad \Longrightarrow \quad x = -(1/2)A^T\nu$$

- plug in in L to obtain g:

$$g(\nu) = L((-1/2)A^T\nu, \nu) = -\frac{1}{4}\nu^T A A^T \nu - b^T\nu$$

a concave function of ν

lower bound property: $p^\star \geq -(1/4)\nu^T A A^T \nu - b^T\nu$ for all ν

Standard form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \quad x \succeq 0 \end{array}$$

dual function

- Lagrangian is

$$\begin{array}{rcl} L(x, \lambda, \nu) & = & c^T x + \nu^T (Ax - b) - \lambda^T x \\ & = & -b^T \nu + (c + A^T \nu - \lambda)^T x \end{array}$$

- L is linear in x, hence

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

  g is linear on affine domain $\{(\lambda, \nu) \mid A^T \nu - \lambda + c = 0\}$, hence concave

lower bound property: $p^\star \geq -b^T \nu$ if $A^T \nu + c \succeq 0$

Equality constrained norm minimization

$$\begin{array}{ll} \text{minimize} & \|x\| \\ \text{subject to} & Ax = b \end{array}$$

dual function

$$g(\nu) = \inf_x (\|x\| - \nu^T Ax + b^T \nu) = \begin{cases} b^T \nu & \|A^T \nu\|_* \leq 1 \\ -\infty & \text{otherwise} \end{cases}$$

where $\|v\|_* = \sup_{\|u\| \leq 1} u^T v$ is dual norm of $\| \cdot \|$

proof: follows from $\inf_x (\|x\| - y^T x) = 0$ if $\|y\|_* \leq 1$, $-\infty$ otherwise

- if $\|y\|_* \leq 1$, then $\|x\| - y^T x \geq 0$ for all x, with equality if $x = 0$
- if $\|y\|_* > 1$, choose $x = tu$ where $\|u\| \leq 1$, $u^T y = \|y\|_* > 1$:

$$\|x\| - y^T x = t(\|u\| - \|y\|_*) \to -\infty \quad \text{as } t \to \infty$$

lower bound property: $p^\star \geq b^T \nu$ if $\|A^T \nu\|_* \leq 1$

Two-way partitioning

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

- a nonconvex problem; feasible set contains 2^n discrete points
- interpretation: partition $\{1, \ldots, n\}$ in two sets; W_{ij} is cost of assigning i, j to the same set; $-W_{ij}$ is cost of assigning to different sets

dual function

$$\begin{aligned} g(\nu) = \inf_x (x^T W x + \sum_i \nu_i (x_i^2 - 1)) &= \inf_x x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu \\ &= \begin{cases} -\mathbf{1}^T \nu & W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

lower bound property: $p^\star \geq -\mathbf{1}^T \nu$ if $W + \mathbf{diag}(\nu) \succeq 0$

example: $\nu = -\lambda_{\min}(W)\mathbf{1}$ gives bound $p^\star \geq n\lambda_{\min}(W)$

The dual problem

Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- finds best lower bound on $p^\star$, obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted $d^\star$
- λ, ν are dual feasible if $\lambda \succeq 0$, $(\lambda, \nu) \in \mathbf{dom}\, g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \mathbf{dom}\, g$ explicit

example: standard form LP and its dual (page 8)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array} \qquad \begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T \nu + c \succeq 0 \end{array}$$

Weak and strong duality

weak duality: $d^\star \leq p^\star$

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems

  for example, solving the SDP

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T \nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0 \end{array}$$

  gives a lower bound for the two-way partitioning problem on page 10

strong duality: $d^\star = p^\star$

- does not hold in general
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called **constraint qualifications**

Slater's constraint qualification

strong duality holds for a convex problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

if it is **strictly feasible**, *i.e.*,

$$\exists x \in \mathbf{int}\, \mathcal{D}: \qquad f_i(x) < 0, \quad i = 1, \ldots, m, \qquad Ax = b$$

- also guarantees that the dual optimum is attained (if $p^\star > -\infty$)
- can be sharpened: *e.g.*, can replace $\mathbf{int}\, \mathcal{D}$ with $\mathbf{relint}\, \mathcal{D}$ (interior relative to affine hull); linear inequalities do not need to hold with strict inequality, . . .
- there exist many other types of constraint qualifications

Feasibility problems

feasibility problem A in $x \in \mathbb{R}^n$.

$$f_i(x) < 0, \quad i = 1, \ldots, m, \qquad h_i(x) = 0, \quad i = 1, \ldots, p$$

feasibility problem B in $\lambda \in \mathbb{R}^m$, $\nu \in \mathbb{R}^p$.

$$\lambda \succeq 0, \qquad \lambda \neq 0, \qquad g(\lambda, \nu) \geq 0$$

where $g(\lambda, \nu) = \inf_x \left(\sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right)$

- feasibility problem B is convex (g is concave), even if problem A is not
- A and B are always **weak alternatives**: at most one is feasible

  proof: assume $\tilde{x}$ satisfies A, λ, ν satisfy B

  $$0 \leq g(\lambda, \nu) \leq \sum_{i=1}^m \lambda_i f_i(\tilde{x}) + \sum_{i=1}^p \nu_i h_i(\tilde{x}) < 0$$

- A and B are **strong alternatives** if exactly one of the two is feasible (can prove infeasibility of A by producing solution of B and vice-versa).

Inequality form LP

primal problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array}$$

dual function

$$g(\lambda) = \inf_x \left((c + A^T\lambda)^T x - b^T\lambda \right) = \left\{ \begin{array}{ll} -b^T\lambda & A^T\lambda + c = 0 \\ -\infty & \text{otherwise} \end{array} \right.$$

dual problem

$$\begin{array}{ll} \text{maximize} & -b^T\lambda \\ \text{subject to} & A^T\lambda + c = 0, \quad \lambda \succeq 0 \end{array}$$

- from Slater's condition: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^\star = d^\star$ except when primal and dual are infeasible

Quadratic program

primal problem (assume $P \in \mathbf{S}^n_{++}$)

$$\begin{array}{ll} \mbox{minimize} & x^TPx \\ \mbox{subject to} & Ax \preceq b \end{array}$$

dual function

$$g(\lambda) = \inf_x \left(x^TPx + \lambda^T(Ax-b)\right) = -\frac{1}{4}\lambda^TAP^{-1}A^T\lambda - b^T\lambda$$

dual problem

$$\begin{array}{ll} \mbox{maximize} & -(1/4)\lambda^TAP^{-1}A^T\lambda - b^T\lambda \\ \mbox{subject to} & \lambda \succeq 0 \end{array}$$

- from Slater's condition: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^\star = d^\star$ always

A nonconvex problem with strong duality

$$\begin{array}{ll} \text{minimize} & x^T A x + 2b^T x \\ \text{subject to} & x^T x \leq 1 \end{array}$$

nonconvex if $A \not\succeq 0$

dual function: $g(\lambda) = \inf_x (x^T(A + \lambda I)x + 2b^T x - \lambda)$

- unbounded below if $A + \lambda I \not\succeq 0$ or if $A + \lambda I \succeq 0$ and $b \notin \mathcal{R}(A + \lambda I)$
- minimized by $x = -(A + \lambda I)^\dagger b$ otherwise: $g(\lambda) = -b^T(A + \lambda I)^\dagger b - \lambda$

dual problem and equivalent SDP:

$$\begin{array}{ll} \text{maximize} & -b^T(A + \lambda I)^\dagger b - \lambda \\ \text{subject to} & A + \lambda I \succeq 0 \\ & b \in \mathcal{R}(A + \lambda I) \end{array} \qquad \begin{array}{ll} \text{maximize} & -t - \lambda \\ \text{subject to} & \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \succeq 0 \end{array}$$

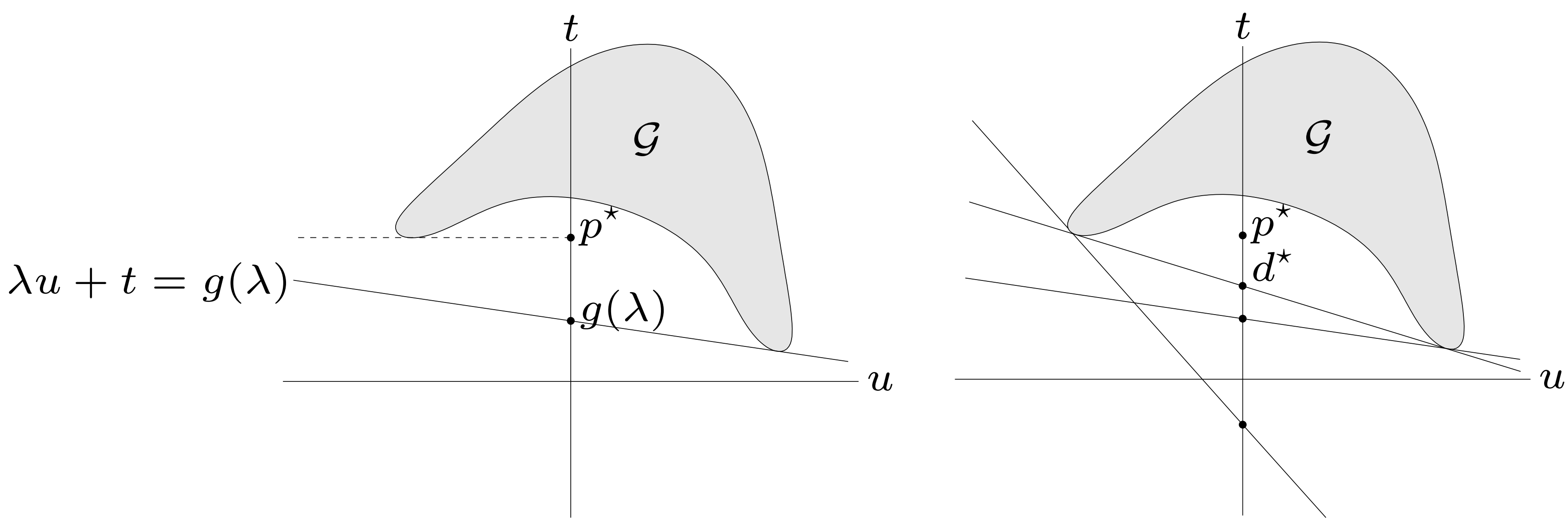
strong duality although primal problem is not convex (not easy to show)

Geometric interpretation

For simplicity, consider problem with one constraint $f_1(x) \leq 0$

interpretation of dual function:

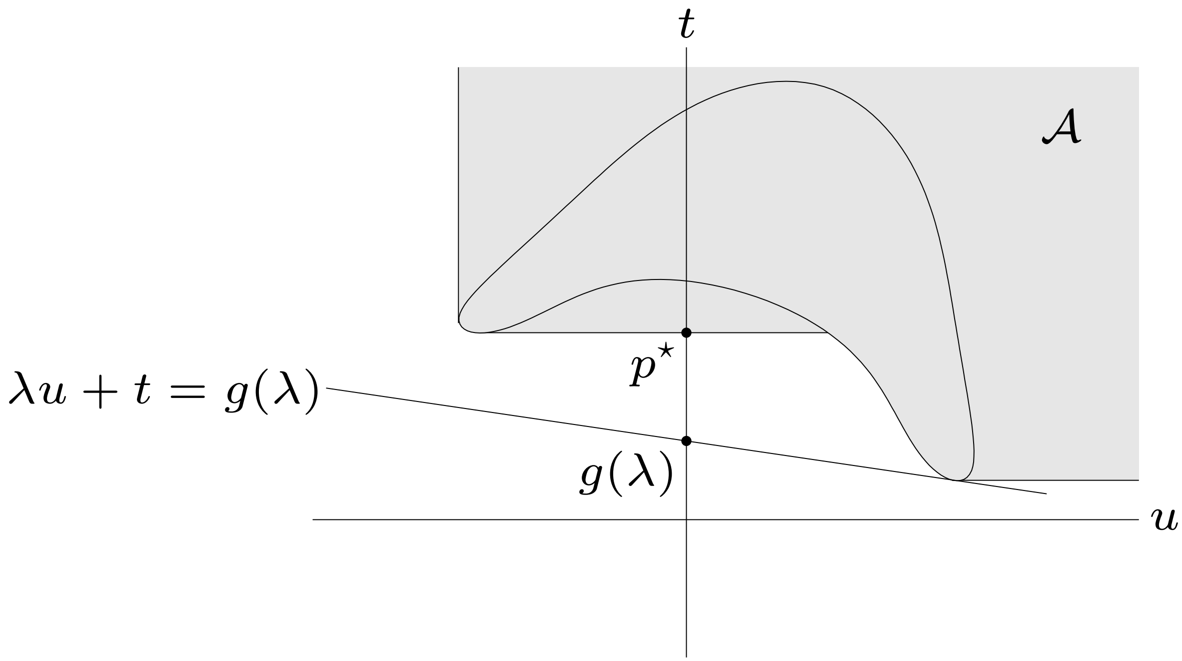
$$g(\lambda) = \inf_{(u,t)\in\mathcal{G}} (t + \lambda u), \qquad \text{where} \quad \mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\}$$



- $\lambda u + t = g(\lambda)$ is (non-vertical) supporting hyperplane to $\mathcal{G}$
- hyperplane intersects t-axis at $t = g(\lambda)$

epigraph variation: same interpretation if $\mathcal{G}$ is replaced with

$$\mathcal{A} = \{(u,t) \mid f_1(x) \leq u, f_0(x) \leq t \text{ for some } x \in \mathcal{D}\}$$



strong duality

- holds if there is a non-vertical supporting hyperplane to $\mathcal{A}$ at $(0, p^\star)$
- for convex problem, $\mathcal{A}$ is convex, hence has supp. hyperplane at $(0, p^\star)$
- Slater's condition: if there exist $(\tilde{u}, \tilde{t}) \in \mathcal{A}$ with $\tilde{u} < 0$, then supporting hyperplanes at $(0, p^\star)$ must be non-vertical

Complementary slackness

Assume strong duality holds, $x^\star$ is primal optimal, $(\lambda^\star, \nu^\star)$ is dual optimal

$$\begin{aligned}
f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\
&\leq f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \\
&\leq f_0(x^\star)
\end{aligned}$$

hence, the two inequalities hold with equality

- $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$
- $\lambda_i^\star f_i(x^\star) = 0$ for $i = 1, \ldots, m$ (known as **complementary slackness**):

$$\lambda_i^\star > 0 \Longrightarrow f_i(x^\star) = 0, \qquad f_i(x^\star) < 0 \Longrightarrow \lambda_i^\star = 0$$

Karush-Kuhn-Tucker (KKT) conditions

the following four conditions are called KKT conditions (for a problem with differentiable f_i, h_i):

1. **Primal feasibility**: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $h_i(x) = 0$, $i = 1, \ldots, p$
2. **Dual feasibility**: $\lambda \succeq 0$
3. **Complementary slackness**: $\lambda_i f_i(x) = 0$, $i = 1, \ldots, m$
4. Gradient of Lagrangian with respect to x vanishes (**first order condition**):

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

If strong duality holds and x, λ, ν are optimal, then they must satisfy the KKT conditions

KKT conditions for convex problem

If $\tilde{x}$, $\tilde{\lambda}$, $\tilde{\nu}$ satisfy KKT for a **convex problem**, then they are optimal:

- from complementary slackness: $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- from 4th condition (and convexity): $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

hence, $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$

If **Slater's condition** is satisfied, x is optimal if and only if there exist λ, ν that satisfy KKT conditions

- recall that Slater implies strong duality, and dual optimum is attained
- generalizes optimality condition $\nabla f_0(x) = 0$ for unconstrained problem

Summary:

- When strong duality holds, the KKT conditions are necessary conditions for optimality
- If the problem is **convex**, they are also sufficient

example: water-filling (assume $\alpha_i > 0$)

$$\begin{array}{ll}\text{minimize} & -\sum_{i=1}^n \log(x_i + \alpha_i) \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1\end{array}$$

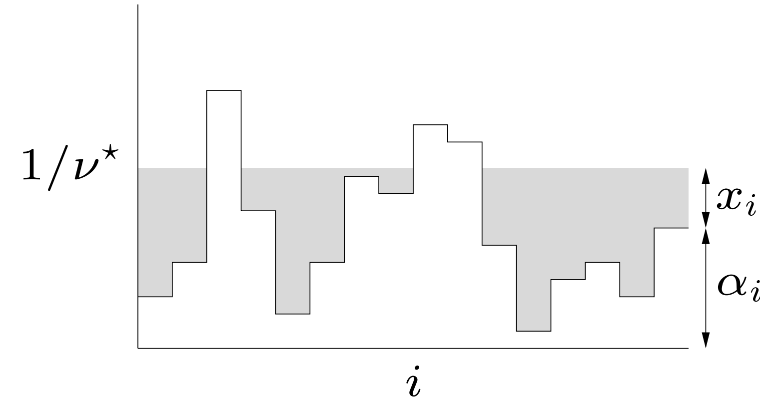
x is optimal iff $x \succeq 0$, $\mathbf{1}^T x = 1$, and there exist $\lambda \in \mathbb{R}^n$, $\nu \in \mathbb{R}$ such that

$$\lambda \succeq 0, \qquad \lambda_i x_i = 0, \qquad \frac{1}{x_i + \alpha_i} + \lambda_i = \nu$$

- if $\nu < 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$
- if $\nu \geq 1/\alpha_i$: $\lambda_i = \nu - 1/\alpha_i$ and $x_i = 0$
- determine ν from $\mathbf{1}^T x = \sum_{i=1}^n \max\{0, 1/\nu - \alpha_i\} = 1$

interpretation

- n patches; level of patch i is at height α_i
- flood area with unit amount of water
- resulting level is $1/\nu^\star$



Perturbation and sensitivity analysis

(unperturbed) optimization problem and its dual

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array} \qquad \begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

perturbed problem and its dual

$$\begin{array}{ll} \text{min.} & f_0(x) \\ \text{s.t.} & f_i(x) \leq u_i, \quad i = 1, \ldots, m \\ & h_i(x) = v_i, \quad i = 1, \ldots, p \end{array} \qquad \begin{array}{ll} \text{max.} & g(\lambda, \nu) - u^T\lambda - v^T\nu \\ \text{s.t.} & \lambda \succeq 0 \end{array}$$

- x is primal variable; u, v are parameters
- $p^\star(u, v)$ is optimal value as a function of u, v
- we are interested in information about $p^\star(u, v)$ that we can obtain from the solution of the unperturbed problem and its dual

Perturbation and sensitivity analysis

global sensitivity result Strong duality holds for unperturbed problem and $\lambda^\star$, $\nu^\star$ are dual optimal for unperturbed problem. Apply **weak duality** to perturbed problem:

$$\begin{aligned} p^\star(u,v) &\geq g(\lambda^\star,\nu^\star) - u^T\lambda^\star - v^T\nu^\star \\ &= p^\star(0,0) - u^T\lambda^\star - v^T\nu^\star \end{aligned}$$

local sensitivity: if (in addition) $p^\star(u,v)$ is differentiable at $(0,0)$, then

$$\lambda_i^\star = -\frac{\partial p^\star(0,0)}{\partial u_i}, \qquad \nu_i^\star = -\frac{\partial p^\star(0,0)}{\partial v_i}$$

Duality and problem reformulations

- equivalent formulations of a problem can lead to very different duals
- reformulating the primal problem can be useful when the dual is difficult to derive, or uninteresting

common reformulations

- introduce new variables and equality constraints
- make explicit constraints implicit or vice-versa
- transform objective or constraint functions

  $e.g.$, replace $f_0(x)$ by $\phi(f_0(x))$ with ϕ convex, increasing

Introducing new variables and equality constraints

$$\text{minimize} \quad f_0(Ax+b)$$

- dual function is constant: $g = \inf_x L(x) = \inf_x f_0(Ax+b) = p^\star$
- we have strong duality, but dual is quite useless

reformulated problem and its dual

$$\begin{array}{ll} \text{minimize} & f_0(y) \\ \text{subject to} & Ax+b-y=0 \end{array} \qquad \begin{array}{ll} \text{maximize} & b^T\nu - f_0^*(\nu) \\ \text{subject to} & A^T\nu = 0 \end{array}$$

dual function follows from

$$\begin{aligned} g(\nu) &= \inf_{x,y}(f_0(y) - \nu^T y + \nu^T Ax + b^T\nu) \\ &= \begin{cases} -f_0^*(\nu) + b^T\nu & A^T\nu = 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

norm approximation problem: minimize $\|Ax - b\|$

$$\begin{array}{ll} \text{minimize} & \|y\| \\ \text{subject to} & y = Ax - b \end{array}$$

can look up conjugate of $\|\cdot\|$, or derive dual directly

$$\begin{aligned} g(\nu) &= \inf_{x,y}(\|y\| + \nu^T y - \nu^T Ax + b^T \nu) \\ &= \begin{cases} b^T\nu + \inf_y(\|y\| + \nu^T y) & A^T\nu = 0 \\ -\infty & \text{otherwise} \end{cases} \\ &= \begin{cases} b^T\nu & A^T\nu = 0, \quad \|\nu\|_* \leq 1 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

(see page 7)

dual of norm approximation problem

$$\begin{array}{ll} \text{maximize} & b^T\nu \\ \text{subject to} & A^T\nu = 0, \quad \|\nu\|_* \leq 1 \end{array}$$

Implicit constraints

LP with box constraints: primal and dual problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & -\mathbf{1} \preceq x \preceq \mathbf{1} \end{array} \qquad \begin{array}{ll} \text{maximize} & -b^T\nu - \mathbf{1}^T\lambda_1 - \mathbf{1}^T\lambda_2 \\ \text{subject to} & c + A^T\nu + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \succeq 0, \quad \lambda_2 \succeq 0 \end{array}$$

reformulation with box constraints made implicit

$$\begin{array}{ll} \text{minimize} & f_0(x) = \begin{cases} c^T x & -\mathbf{1} \preceq x \preceq \mathbf{1} \\ \infty & \text{otherwise} \end{cases} \\ \text{subject to} & Ax = b \end{array}$$

dual function

$$\begin{aligned} g(\nu) &= \inf_{-\mathbf{1} \preceq x \preceq \mathbf{1}} (c^T x + \nu^T(Ax - b)) \\ &= -b^T\nu - \|A^T\nu + c\|_1 \end{aligned}$$

dual problem: maximize $-b^T\nu - \|A^T\nu + c\|_1$

Problems with generalized inequalities

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \preceq_{K_i} 0, \quad i=1,\ldots,m \\ & h_i(x) = 0, \quad i=1,\ldots,p \end{array}$$

$\preceq_{K_i}$ is generalized inequality on $\mathbb{R}^{k_i}$

definitions are parallel to scalar case:

- Lagrange multiplier for $f_i(x) \preceq_{K_i} 0$ is vector $\lambda_i \in \mathbb{R}^{k_i}$
- Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_m} \times \mathbb{R}^p \to \mathbb{R}$, is defined as

$$L(x, \lambda_1, \cdots, \lambda_m, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i^T f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- dual function $g : \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_m} \times \mathbb{R}^p \to \mathbb{R}$, is defined as

$$g(\lambda_1, \ldots, \lambda_m, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda_1, \cdots, \lambda_m, \nu)$$

lower bound property: if $\lambda_i \succeq_{K_i^*} 0$, then $g(\lambda_1, \ldots, \lambda_m, \nu) \leq p^\star$

proof: if $\tilde{x}$ is feasible and $\lambda \succeq_{K_i^*} 0$, then

$$\begin{aligned} f_0(\tilde{x}) &\geq f_0(\tilde{x}) + \sum_{i=1}^m \lambda_i^T f_i(\tilde{x}) + \sum_{i=1}^p \nu_i h_i(\tilde{x}) \\ &\geq \inf_{x \in \mathcal{D}} L(x, \lambda_1, \ldots, \lambda_m, \nu) \\ &= g(\lambda_1, \ldots, \lambda_m, \nu) \end{aligned}$$

minimizing over all feasible $\tilde{x}$ gives $p^\star \geq g(\lambda_1, \ldots, \lambda_m, \nu)$

dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda_1, \ldots, \lambda_m, \nu) \\ \text{subject to} & \lambda_i \succeq_{K_i^*} 0, \quad i = 1, \ldots, m \end{array}$$

- weak duality: $p^\star \geq d^\star$ always
- strong duality: $p^\star = d^\star$ for convex problem with constraint qualification (for example, Slater's: primal problem is strictly feasible)

Semidefinite program

primal SDP $(F_i, G \in \mathbf{S}^k)$

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + \cdots + x_n F_n \preceq G \end{array}$$

- Lagrange multiplier is matrix $Z \in \mathbf{S}^k$
- Lagrangian $L(x, Z) = c^T x + \mathbf{Tr}\left(Z(x_1 F_1 + \cdots + x_n F_n - G)\right)$
- dual function

$$g(Z) = \inf_x L(x, Z) = \begin{cases} -\mathbf{Tr}(GZ) & \mathbf{Tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \\ -\infty & \text{otherwise} \end{cases}$$

dual SDP

$$\begin{array}{ll} \text{maximize} & -\mathbf{Tr}(GZ) \\ \text{subject to} & Z \succeq 0, \quad \mathbf{Tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \end{array}$$

$p^\star = d^\star$ if primal SDP is strictly feasible ($\exists x$ with $x_1 F_1 + \cdots + x_n F_n \prec G$)

Duality: SOCP example

Let's consider the following Second Order Cone Program (**SOCP**):

$$\begin{array}{ll}\text{minimize} & f^T x\\ \text{subject to} & \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m,\end{array}$$

with variable $x \in \mathbb{R}^n$. Let's show that the dual can be expressed as

$$\begin{array}{ll}\text{maximize} & \sum_{i=1}^m (b_i^T u_i + d_i v_i)\\ \text{subject to} & \sum_{i=1}^m (A_i^T u_i + c_i v_i) + f = 0\\ & \|u_i\|_2 \leq v_i, \quad i = 1, \ldots, m,\end{array}$$

with variables $u_i \in \mathbb{R}^{n_i}$, $v_i \in \mathbb{R}$, $i = 1, \ldots, m$ and problem data given by $f \in \mathbb{R}^n$, $A_i \in \mathbb{R}^{n_i \times n}$, $b_i \in \mathbb{R}^{n_i}$, $c_i \in \mathbb{R}$ and $d_i \in \mathbb{R}$.

Duality: SOCP

We can derive the dual in the following two ways:

1. Introduce new variables $y_i \in \mathbb{R}^{n_i}$ and $t_i \in \mathbb{R}$ and equalities $y_i = A_i x + b_i$, $t_i = c_i^T x + d_i$, and derive the Lagrange dual.

2. Start from the conic formulation of the SOCP and use the conic dual. Use the fact that the second-order cone is **self-dual**:

$$t \geq \|x\| \iff tv + x^T y \geq 0, \text{ for all } v, y \text{ such that } v \geq \|y\|$$

   The condition $x^T y \leq tv$ is a simple Cauchy-Schwarz inequality

Duality: SOCP

We introduce new variables, and write the problem as

$$\begin{array}{ll}\mbox{minimize} & c^Tx\\ \mbox{subject to} & \|y_i\|_2 \leq t_i, \quad i=1,\ldots,m\\ & y_i = A_ix+b_i,\; t_i=c_i^Tx+d_i, \quad i=1,\ldots,m\end{array}$$

The **Lagrangian** is

$$\begin{array}{rl} & L(x,y,t,\lambda,\nu,\mu)\\ = & c^Tx+\sum_{i=1}^m \lambda_i(\|y_i\|_2-t_i)+\sum_{i=1}^m \nu_i^T(y_i-A_ix-b_i)+\sum_{i=1}^m \mu_i(t_i-c_i^Tx-d_i)\\ = & (c-\sum_{i=1}^m A_i^T\nu_i-\sum_{i=1}^m \mu_ic_i)^Tx+\sum_{i=1}^m(\lambda_i\|y_i\|_2+\nu_i^Ty_i)+\sum_{i=1}^m(-\lambda_i+\mu_i)t_i\\ & -\sum_{i=1}^n(b_i^T\nu_i+d_i\mu_i).\end{array}$$

Duality: SOCP

The minimum over x is bounded below if and only if

$$\sum_{i=1}^{m}(A_i^T\nu_i + \mu_i c_i) = c.$$

To minimize over y_i, we note that

$$\inf_{y_i}(\lambda_i\|y_i\|_2 + \nu_i^T y_i) = \begin{cases} 0 & \|\nu_i\|_2 \leq \lambda_i \\ -\infty & \text{otherwise.} \end{cases}$$

The minimum over t_i is bounded below if and only if $\lambda_i = \mu_i$.

Duality: SOCP

The Lagrange dual function is

$$g(\lambda,\nu,\mu)=\begin{cases}-\sum_{i=1}^n(b_i^T\nu_i+d_i\mu_i) & \text{if } \sum_{i=1}^m(A_i^T\nu_i+\mu_ic_i)=c,\\ & \|\nu_i\|_2\leq\lambda_i,\quad \mu=\lambda\\ -\infty & \text{otherwise}\end{cases}$$

which leads to the dual problem

$$\begin{array}{ll}\text{maximize} & -\sum_{i=1}^n(b_i^T\nu_i+d_i\lambda_i)\\ \text{subject to} & \sum_{i=1}^m(A_i^T\nu_i+\lambda_ic_i)=c\\ & \|\nu_i\|_2\leq\lambda_i,\quad i=1,\ldots,m.\end{array}$$

which is again an SOCP

Duality: SOCP

We can also express the SOCP as a **conic form** problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & -(c_i^T x + d_i, A_i x + b_i) \preceq_{K_i} 0, \quad i = 1, \ldots, m. \end{array}$$

The Lagrangian is given by:

$$\begin{aligned} L(x, u_i, v_i) &= c^T x - \textstyle\sum_i (A_i x + b_i)^T u_i - \sum_i (c_i^T x + d_i) v_i \\ &= \left(c - \textstyle\sum_i (A_i^T u_i + c_i v_i)\right)^T x - \sum_i (b_i^T u_i + d_i v_i) \end{aligned}$$

for $(v_i, u_i) \succeq_{K_i^*} 0$ (which is also $v_i \geq \|u_i\|$)

Duality: SOCP

With

$$L(x,u_i,v_i) = \left(c - \sum_i (A_i^T u_i + c_i v_i)\right)^T x - \sum_i (b_i^T u_i + d_i v_i)$$

the **dual function** is given by:

$$g(\lambda,\nu,\mu) = \begin{cases} -\sum_{i=1}^n (b_i^T \nu_i + d_i \mu_i) & \text{if } \sum_{i=1}^m (A_i^T \nu_i + \mu_i c_i) = c, \\ -\infty & \text{otherwise} \end{cases}$$

The **conic dual** is then:

$$\begin{array}{ll} \text{maximize} & -\sum_{i=1}^n (b_i^T u_i + d_i v_i) \\ \text{subject to} & \sum_{i=1}^m (A_i^T u_i + v_i c_i) = c \\ & (v_i, u_i) \succeq_{K_i^*} 0, \quad i = 1,\ldots,m. \end{array}$$

Proof

Convex problem & constraint qualification

$$\Downarrow$$

Strong duality

Slater's constraint qualification

Convex problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

The problem satisfies Slater's condition if it is **strictly feasible**, $i.e.$,

$$\exists x \in \mathbf{int}\, \mathcal{D}: \qquad f_i(x) < 0, \quad i = 1, \ldots, m, \qquad Ax = b$$

- also guarantees that the dual optimum is attained (if $p^\star > -\infty$)
- there exist many other types of constraint qualifications

KKT conditions for convex problem

If $\tilde{x}$, $\tilde{\lambda}$, $\tilde{\nu}$ satisfy KKT for a **convex problem**, then they are optimal:

- from complementary slackness: $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- from 4th condition (and convexity): $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

hence, $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$ with $(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$ feasible.

If **Slater's condition** is satisfied, x is optimal if and only if there exist λ, ν that satisfy KKT conditions

- Slater implies strong duality (more on this now), and dual optimum is attained
- generalizes optimality condition $\nabla f_0(x) = 0$ for unconstrained problem

Summary

- For a convex problem satisfying constraint qualification, the KKT conditions are **necessary & sufficient** conditions for optimality.

Proof

To simplify the analysis. We make two additional technical assumptions:

- The domain $\mathcal{D}$ has nonempty interior (hence, $\mathbf{relint}\,\mathcal{D} = \mathbf{int}\,\mathcal{D}$)
- We also assume that A has full rank, i.e. $\mathbf{Rank}\,A = p$.

Proof

- We define the set $\mathcal{A}$ as

$$\mathcal{A} = \{(u, v, t) \mid \exists x \in \mathcal{D},\ f_i(x) \leq u_i,\ i = 1, \ldots, m,\\ h_i(x) = v_i,\ i = 1, \ldots, p,\ f_0(x) \leq t\},$$

  which is the set of values taken by the constraint and objective functions.

- If the problem is convex, $\mathcal{A}$ is defined by a list of convex constraints hence is **convex**.

- We define a second convex set $\mathcal{B}$ as

$$\mathcal{B} = \{(0, 0, s) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \mid s < p^\star\}.$$

- The sets $\mathcal{A}$ and $\mathcal{B}$ do not intersect (otherwise $p^\star$ could not be optimal value of the problem).

First step: The hyperplane separating $\mathcal{A}$ and $\mathcal{B}$ defines a supporting hyperplane to $\mathcal{A}$ at $(0, p^\star)$.

Geometric proof

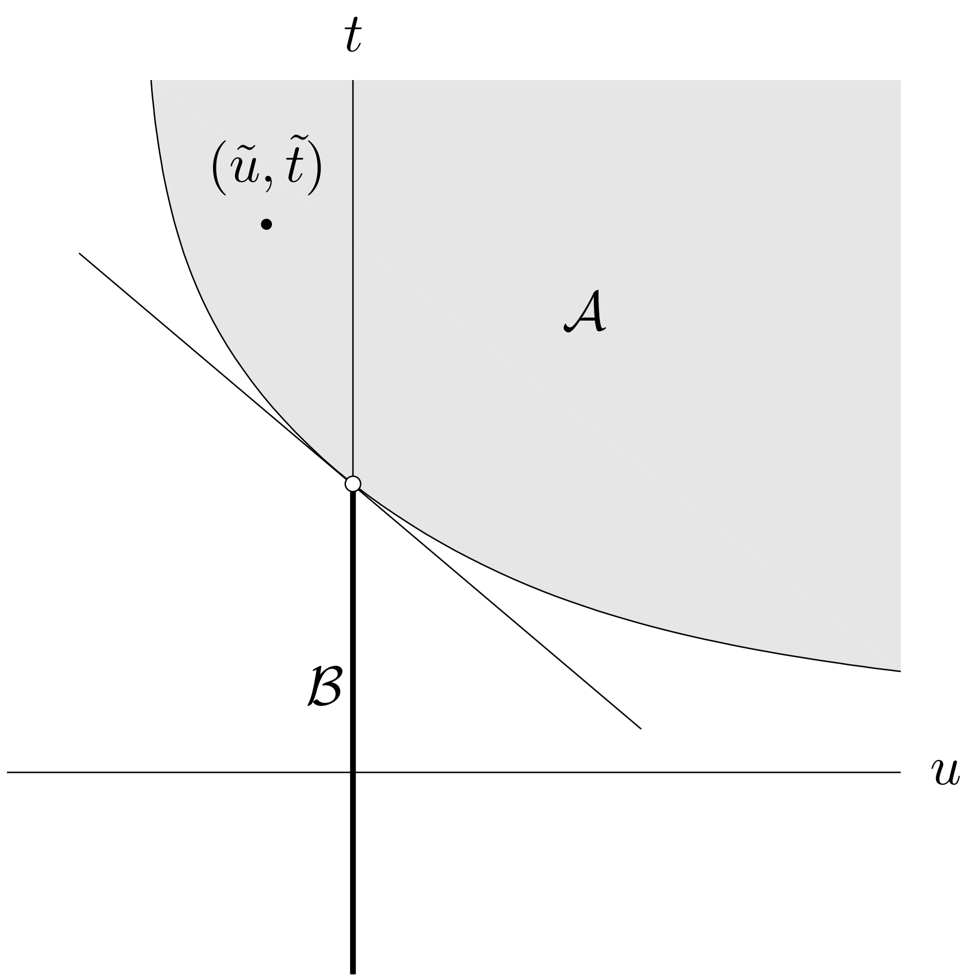


Illustration of strong duality proof, for a convex problem that satisfies Slater's constraint qualification. The two sets $\mathcal{A}$ and $\mathcal{B}$ are convex and do not intersect, so they can be separated by a hyperplane. Slater's constraint qualification guarantees that any separating hyperplane must be nonvertical.

Proof

- By the separating hyperplane theorem there exists $(\tilde{\lambda}, \tilde{\nu}, \mu) \neq 0$ and α such that
$$(u, v, t) \in \mathcal{A} \implies \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \geq \alpha, \tag{1}$$
and
$$(u, v, t) \in \mathcal{B} \implies \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \leq \alpha. \tag{2}$$
- From (1) we conclude that $\tilde{\lambda} \succeq 0$ and $\mu \geq 0$. (Otherwise $\tilde{\lambda}^T u + \mu t$ is unbounded below over $\mathcal{A}$, contradicting (1).)
- The condition (2) simply means that $\mu t \leq \alpha$ for all $t < p^\star$, and hence, $\mu p^\star \leq \alpha$.

Together with (1) we conclude that for any $x \in \mathcal{D}$,

$$\mu p^\star \leq \alpha \leq \mu f_0(x) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T (Ax - b) \tag{3}$$

Proof

Let us assume that $\mu > 0$ (separating hyperplane is nonvertical)

- We can divide the previous equation by μ to get

$$L(x, \tilde{\lambda}/\mu, \tilde{\nu}/\mu) \geq p^\star$$

  for all $x \in \mathcal{D}$

- Minimizing this inequality over x produces $p^\star \leq g(\lambda, \nu)$, where

$$\lambda = \tilde{\lambda}/\mu, \qquad \nu = \tilde{\nu}/\mu.$$

- By weak duality we have $g(\lambda, \nu) \leq p^\star$, so in fact $g(\lambda, \nu) = p^\star$.

This shows that strong duality holds, and that the dual optimum is attained, whenever $\mu > 0$. The normal vector has the form $(\lambda^\star, 1)$ and produces the Lagrange multipliers.

Proof

Second step: Slater's constraint qualification is used to establish that the hyperplane must be **nonvertical**, i.e. $\mu > 0$.

By contradiction, assume that $\mu = 0$. From (3), we conclude that for all $x \in \mathcal{D}$,

$$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T (Ax - b) \geq 0. \tag{4}$$

- Applying this to the point $\tilde{x}$ that satisfies the Slater condition, we have

$$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x}) \geq 0.$$

- Since $f_i(\tilde{x}) < 0$ and $\tilde{\lambda}_i \geq 0$, we conclude that $\tilde{\lambda} = 0$.

Proof

This is where we use the two technical assumptions.

- Then (4) implies that for all $x \in \mathcal{D}$, $\tilde{\nu}^T(Ax - b) \geq 0$.
- But $\tilde{x}$ satisfies $\tilde{\nu}^T(A\tilde{x} - b) = 0$, and since $\tilde{x} \in \mathbf{int}\, \mathcal{D}$, there are points in $\mathcal{D}$ with $\tilde{\nu}^T(Ax - b) < 0$ unless $A^T\tilde{\nu} = 0$.
- This contradicts our assumption that $\mathbf{Rank}\, A = p$.

This means that we cannot have $\mu = 0$ and ends the proof.